NAMA: DITO DWI PRASETYO

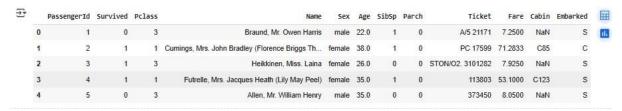
KELAS: 3J TEKNIK INFORMATIKA

NPM : 2213020043

```
[1] import pandas as pd
  import numpy as np
  import seaborn as sns
  import matplotlib.pyplot as plt
  import missingno as msno

[9] df = pd.read_csv('Titanic-Dataset.csv') #Ambil Data
  df.shape #Menampilkan Ukuran Data

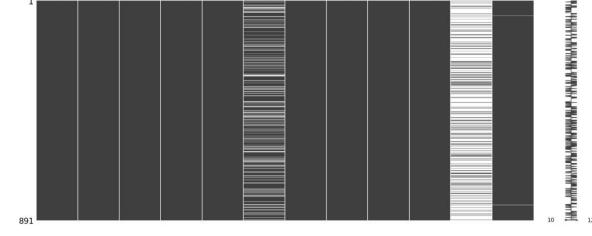
  df.head() #Menampilkan 5 Baris Data
```



```
[10] print("\n--- Informasi Data ---")
    df.info() #Informasi Dari Dataset
```

```
∓÷
    --- Informasi Data ---
    <class 'pandas.core.frame.DataFrame'>
    RangeIndex: 891 entries, 0 to 890
    Data columns (total 12 columns):
    # Column
                   Non-Null Count Dtype
                    .....
    0 PassengerId 891 non-null int64
    1 Survived 891 non-null
                                 int64
int64
    2
        Pclass
                    891 non-null
                 891 non-null
       Name
                                 object
                  891 non-null
    4 Sex
                                 object
       Age
SibSp
    5
                    714 non-null
                                  float64
                  714 non ....
891 non-null
    6
                                  int64
                  891 non-null
891 non-null
       Parch
                                 int64
    8 Ticket
                                 object
    9 Fare
10 Cabin
                    891 non-null
                                  float64
                   204 non-null
                                  object
    11 Embarked
                  889 non-null
                                 object
    dtypes: float64(2), int64(5), object(5)
    memory usage: 83.7+ KB
```

```
[11] print("\n--- Missing Values ---")
      print(df.isnull().sum()) #Menampilkan Jumlah Missing Values
₹
      --- Missing Values ---
     PassengerId
                        0
     Survived
                        0
     Pclass.
                        0
     Name
                        0
     Sex
     Age
                      177
     SibSp
                        0
     Parch
      Ticket
                        0
      Fare
                        0
     Cabin
                      687
     Embarked
     dtype: int64
[12] #Visualisasi Missing Values
     plt.figure(figsize=(10,5))
     msno.matrix(df)
     plt.title("Visualisasi Missing Values")
     plt.show()
                                               Visualisasi Missing Values
                                                        SIDSP
                                                                Parch
                                                                        Ticket
                                                 poe
```



Berdasarkan dari dataset "Titanic-Dataset.csv" terdapat 3 atribut yang memiliki missing value (data yang tidak memiliki nilai / kosong), sebagai berikut :

- 1. Age (177 missing values)
 - Tipe Data : Numerik (usia penumpang).
 - Cara Menanganinya:
 - Mengisi dengan median usia agar tidak terlalu terpengaruh oleh outlier.
 - Mengelompokkan berdasarkan kelas tiket (Pclass) dan mengisi dengan median usia dari tiap kelas (karena usia dan kelas memiliki korelasi).
 - Alasannya data tidak diisi oleh penumpang atau tidak dicatat dengan baik.

2. Cabin (687 missing values)

- Tipe Data: Kategorikal (nomor kabin).
- Cara Menanganinya:
 - Dikarenakan terlalu banyak data yang hilang, lebih baik menghapus kolom ini jika tidak terlalu penting.
 - Mengelompokkan berdasarkan kelas tiket (Pclass) dan memberikan label kategori seperti "Unknown" untuk penumpang tanpa kabin.
- Alasannya banyak penumpang tidak memiliki informasi kabin, terutama dari kelas ekonomi.

3. Embarked (2 missing values)

- Tipe Data: Kategorikal (pelabuhan tempat naik: C = Cherbourg, Q = Queenstown, S = Southampton).
- Cara Menangani:
 - Karena hanya 2 data yang hilang, bisa diisi dengan modus (nilai yang paling sering muncul), yaitu Southampton (S).
- Alasannya Data tidak tercatat atau hilang.